AFRL-SR-AR-TR-03-
0174

REPORT DOCUMENTATION PAGE

1. Agency Use Only (Leave blank)	2. Report Date	3. Report Type and Period Covered (i.e., annual 1 Jun 00 - 31 May 01)
	January 21, 2003	Final Report for Period 1 Dec 99 to 30 Nov 02

4. Title and Subtitle

Data Mining via Generalized Support Vector Machines

5. Award Number

F49620-00-1-0085

6. Author(s)

Olvi L. Mangasarian

7. Performing Organization Name (Include Name, City, State, Zip Code and Email for Principal Investigator)

University of Wisconsin
Madison, WI 53706

E-Mail: olvi@cs.wisc.edu

8. Performing Organization Report Number (Leave Blank)

9. Sponsoring/Monitoring Agency Name and Address

Air Force Office of Scientific Research
4015 Wilson Blvd., Room 713
Arlington, VA 22203-1954

10. Sponsoring/Monitoring Agency Report Number (Leave Blank)

11. Supplementary Notes (i.e., report contains color photos, report contains appendix in non-print form, etc.)

12a. Distribution/Availability Statement (check one)

X Approved for public release; distribution unlimited

☐ ~~Distribution limited to U.S. Government agencies only - report contains proprietary information~~

12b. Distribution Code (Leave Blank)

13. Abstract *(Maximum 200 Words)* (abstract should contain no proprietary or confidential information)

Generalized Support Vector Machines were used to extract valuable information from datasets and construct fast classification algorithms for massive data. The influence of chemotherapy was investigated on breast cancer patients by obtaining well separated Kaplan-Meier survival curves for three classes of patients. A novel approach was proposed for using a minimal number of data points in order to generate an accurate classifier. Substantial progress was also made towards achieving new results in the field of data mining by using the extremely versatile and highly effective approach of support vector machines. In particular minimal kernel classifiers were constructed that use a minimal subset of the data. A new type of classifier, the proximal classifier, was proposed and implemented which is typically an order of magnitude faster than conventional classifiers. The effect of chemotherapy on breast cancer patients was more accurately assessed. An incremental classification algorithm was proposed, implemented and was capable of classifying a billion points in less than three hours on a 400Mhz machine. New techniques for incorporating prior expert knowledge, such as medical doctors' experience, into classifiers were devised and computationally implemented. Very fast Newton methods were proposed and successfully tested for extremely large classification problems and linear programming problems.

20030520 114

14. Subject Terms (keywords previously assigned to proposal abstract or terms which apply to this award) Data mining, classification, support vector machines, medical applications, optimization			15. Number of Pages (count all pages including appendices) 13
			16. Price Code (Leave Blank)
17. Security Classification of Report Unclassified	18. Security Classification of this Page Unclassified	19. Security Classification of Abstract Unclassified	20. Limitation of Abstract Unlimited

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. Z39-18
298-102

1. Final Report on AFOSR Grant F49620-00-1-0085
"Data Mining via Generalized Support Vector Machines"

PI: Olvi L. Mangasarian olvi@cs.wisc.edu
www.cs.wisc.edu/~olvi
University of Wisconsin - Madison

Reporting Period: 1 Dec 99 to 30 Nov 02

2. Objectives

Use a very general formulation of Support Vector Machines to extract significant information and knowledge from datasets.

3. Abstract of Results

During the first year of the grant, Generalized Support Vector Machines were used to extract valuable information from datasets and construct fast classification algorithms for massive data. The influence of chemotherapy was investigated on breast cancer patients by obtaining well separated Kaplan-Meier survival curves for three classes of patients. A novel approach was proposed for using a minimal number of data points in order to generate an accurate classifier. Simple and powerful iterative methods for classifying massive datasets have been given, one of which requires only 11 lines of MATLAb code. A method has been proposed for generating complex nonlinear classifiers based on as little as 1% of the given data.

During the seocnd year of the grant, substantial progress was made towards achieving new results in the field of data mining by using the extremely versatile and highly effective approach of support vector machines. In particular minimal kernel classifiers were constructed that use a minimal subset of the data, a new type of classifier, the proximal classifier, was proposed and implemented which is typically an order of magnitude faster than conventional classifiers, the effect of chemotherapy on breast cancer patients was more accurately assessed than before and a new fast multicategory classifier was proposed and implemented.

During the third and final year of the grant, significant progress was made in a number of research directions under this project. An incremental classification algorithm was proposed and implemented which was capable of classifying a billion points in less than three hours on a 400Mhz machine. New techniques for incorporating prior expert knowledge, such as medical doctors' experience, into classifiers were devised and computationally implemented. This work led to new mathematical theoretical results on set containment problems. Very fast Newton methods were proposed and successfully tested for extremely large classification problems and linear programming problems.

4. Accomplishments/New Findings

The research supported by this grant resulted in:

(a) Seventeen papers, 10 of which are already published, 3 have been accepted, and 4 have been submitted, all to journals or national or international conferences. These papers are listed in Section 6 and are easily available on the web as indicated by the links given in Section 6 following each paper, or from the PI's home page: www.cs.wisc.edu/~olvi

(b) Seventeen talks given at national and international conferences, workshops and at universities.

4a. Summary of Results (Numbers refer to Section 6 below)

In (6(i)) a linear support vector machine (SVM) is used to extract 6 features from a total of 31 features in a dataset of 253 breast cancer patients. Five features are nuclear features obtained during a non-invasive diagnostic procedure while one feature, tumor size, is obtained during surgery. The linear SVM selected the 6 features in the process of classifying the patients into node-positive (patients with some metastasized lymph nodes) and node-negative (patients with no metastasized lymph nodes). Node-positive patients are typically those who undergo chemotherapy. The 6 features were then used in a Gaussian kernel nonlinear SVM to classify the patients into three prognostic groups: good (node-negative), intermediate (1 to 4 metastasized nodes) and poor (more than 4 metastasized nodes). Very well separated Kaplan-Meier survival curves were constructed for the three groups with pairwise p-value of less than 0.009 based on the logrank statistic. Patients in the good prognostic group had the highest survival, while patients in the poor prognostic group had the lowest. The majority (72.8%) of the good group did not receive chemotherapy, while the majority (87.5%) of the poor group received chemotherapy. Just over half (56.7%) of the intermediate group received chemotherapy. New patients can be assigned to one of these three prognostic groups with its associated survival curve, based only on 6 features obtained before and during surgery, but without the potentially risky procedure of removing lymph nodes to determine how many of them have metastasized.

In (6(ii)) the problem of extracting a minimal number of data points from a large dataset, in order to generate a support vector machine (SVM) classifier, is formulated as a concave minimization problem and solved by a finite number of linear programs. The minimal set of selected support vectors is considerably smaller than that required by a standard 1-norm support vector machine with or without feature selection. The proposed approach also incorporates a feature selection procedure that results in a minimal number of input features used by the classifier. Tenfold cross validation gives as good or better test results using the proposed minimal support vector machine (MSVM) classifier based on the smaller set of data points compared to a standard 1-norm support vector machine classifier. The reduction in data points used by an MSVM classifier over those used by a 1-norm SVM classifier averaged 66% on seven public datasets and was as high as 81%. This makes MSVM a useful incremental classification tool which maintains only a small fraction of a large dataset before merging and processing it with new incoming data.

In (6(iii)) an active set strategy is applied to the dual of a simple reformulation of the standard quadratic program associated with a linear support vector machine. This application generates a new dual algorithm that consists of solving a finite succession of linear equations for support vector multipliers from which a separating surface is uniquely determined. The linear equations are in the dual space, with a typically large dimensionality equal to the number of points to be classified. However, by making novel use of the Sherman-Morrison-Woodbury formula, a much smaller matrix of the order of the original input space is inverted at each step. Thus, a problem with a 32-dimensional input space and 7 million points required 4 iterations to solve, each iteration consisting of inverting a 33-by-33 positive definite symmetric matrix with a total running time of 73.2 minutes on a 400 MHz Intel Pentium II processor. The proposed algorithm requires no specialized quadratic or linear programming code, but merely a linear equation solver which is publicly available.

In (6(iv)) an implicit Lagrangian for the dual of a simple reformulation of the standard quadratic program of a linear support vector machine is proposed. This leads to the minimization of an UNCONSTRAINED differentiable convex function in a space of dimensionality equal to the number of classified points. This problem is solvable by an extremely simple linearly convergent Lagrangian support vector machine (LSVM) algorithm. LSVM requires the inversion at the outset of a single matrix of the order of the much smaller dimensionality of the original input space plus one. The full algorithm is given in this paper in 11 lines of MATLAB code without any special optimization tools such as linear or quadratic programming solvers. This LSVM code can be used "as is" to solve classification problems with millions of points. For example, 2 million points in 10 dimensional input space were classified in 82 minutes on a Pentium III 500 MHz notebook with 384 megabytes of memory (and additional swap space), and in 7 minutes on a 250 MHz UltraSPARC II processor with 2 gigabytes of memory. Other standard classification test problems were also solved. Nonlinear kernel classification can also be solved by LSVM. Although it does not scale up to very large problems, it can handle any positive semidefinite kernel and is guaranteed to converge. A short MATLAB code is also given for nonlinear kernels and tested on a number of problems.

In (6(v)) an algorithm is proposed which generates a nonlinear kernel-based separating surface that requires as little as 1% of a large dataset for its explicit evaluation. Although the entire dataset is used to generate the nonlinear surface, the final kernel makes explicit use of a very small portion of the data, the remainder of which can be thrown away. This is achieved by making use of a RECTANGULAR m-by-k kernel K(A,B') that greatly reduces the size of the quadratic program to be solved and simplifies the characterization of the nonlinear separating surface. Here, the m rows of A represent the original m data points while the k rows of B represent a greatly reduced k data points. Computational results indicate that test set correctness for the reduced support vector machine (RSVM), with a nonlinear separating surface that depends on a small randomly selected portion of the dataset, is better or much better than that of a conventional support vector machine with a nonlinear surface that explicitly depends on either the entire dataset or the smaller randomly selected one.

In (6(vi)) a finite concave minimization algorithm is proposed for constructing kernel classifiers that use a minimal number of data points both in generating and characterizing a classifier. The algorithm is theoretically justified on the basis of linear programming perturbation theory and a leave-one-out error bound as well as effective computational results on seven real world datasets. A nonlinear rectangular kernel is generated by systematically utilizing as few of the data as possible both in training and in characterizing a nonlinear separating surface. This can result in substantial kernel size reduction (over 87% in six of the seven public datasets tested on) with test set correctness equal to that obtained by using a full kernel. To eliminate data points, the proposed approach makes use of a novel loss function, the ``pound" function (.)#, which is a linear combination of the 1-norm and the step function that measures both the magnitude and the presence of any error. The effectiveness of this loss function is justified by appealing to Bayesian and statistical learning theory.

In (6(vii)) instead of a standard support vector machine (SVM) that classifies points by assigning them to one of two disjoint halfspaces, points are classified by assigning them to the closest of two parallel planes (in input or feature space) that are pushed apart as far as possible. This formulation, which can also be interpreted as regularized least squares, leads to an extremely fast and simple algorithm for generating a linear or nonlinear classifier that merely requires the solution of a single system of linear equations. In contrast, standard SVMs solve a quadratic or a linear program that require considerably longer computational time. Computational results on publicly available datasets indicate that the proposed proximal SVM classifier has comparable test set correctness to that of standard SVM classifiers, but with considerably faster computational time that can be an order of magnitude faster. The linear proximal SVM can easily handle large datasets as indicated by the classification of a 2 million point 10-attribute set in 20.8 seconds.

All computational results are based on 6 lines of MATLAB code.

In (6(viii)) the identification of breast cancer patients for whom chemotherapy could prolong survival time is treated here as a data mining problem. This identification is achieved by clustering 253 breast cancer patients into three prognostic groups: Good, Poor and Intermediate. Each of the three groups has a significantly distinct Kaplan-Meier survival curve. Of particular significance is the Intermediate group, because patients with chemotherapy in this group do better than those without chemotherapy in the same group. This is the reverse case to that of the overall population of 253 patients for which patients undergoing chemotherapy have worse survival than those who do not. We also prescribe a procedure that utilizes three nonlinear smooth support vector machines (SSVMs) for classifying breast cancer patients into the three above prognostic groups. These results suggest that the patients in the Good group should not receive chemotherapy while those in the Intermediate group should receive chemotherapy based on our survival curve analysis. To our knowledge this is the first instance of a classifiable group of breast cancer patients for which chemotherapy can possibly enhance survival.

In (6(ix)) we show how support vector machines (SVMs) have played a key role In broad classes of problems arising in various fields. Much more recently, SVMs have become the tool of choice for problems arising in data classification and mining. This paper emphasizes some recent developments that the author and his colleagues have contributed to such as: generalized SVMs (a very general mathematical programming framework for SVMs), smooth SVMs (a smooth nonlinear equation representation of SVMs solvable by a fast Newton method), Lagrangian SVMs (an unconstrained Lagrangian representation of SVMs leading to an extremely simple iterative scheme capable of solving classification problems with millions of points) and reduced SVMs (a rectangular kernel classifier that utilizes as little as 1% of the data).

In (6(x)) we construct by an extremely fast algorithm, a proximal support vector machine k-category classifier with the following properties:

[(a)] The classifier consists of k linear or nonlinear surfaces in the n-dimensional input space. Each surface separates one category from the rest.
[(b)] The k surfaces are proximal to each of the k categories and are constructed in time that can be as little as 3 orders of magnitude shorter than that of a conventional support vector machine.
[(c)] Each of the k linear or nonlinear surfaces is the unique unconstrained minimizer of a strongly convex quadratic function.
[(d)] Each of the k linear or nonlinear surfaces can be obtained by solving a single nonsingular system of t linear equations in as many unknowns. For linear surfaces, t=(n+1) where n is the dimensionality of the input space in which the dataset resides. For nonlinear surfaces t can be

as small as 15% of the total number of points in the dataset.
[(e)] No quadratic or linear programming codes are needed. Effective linear equation solvers are publicly available and can efficiently solve large problems.

Computational results arc presented which indicate that the method is orders of magnitude faster, but has similar test set correctness to methods using conventional support vector machines that also separate each category from the rest in a consecutive manner by placing the points to be separated in disjoint halfspaces in the input or feature space.

In (6(xi)), using a recently introduced proximal support vector machine classifier by the authors, a very fast and simple incremental support vector machine (SVM) classifier is proposed which is capable of modifying an existing linear classifier by both RETIRING old data and ADDING new data. A very important feature of the proposed single-pass algorithm, which allows it to handle massive datasets, is that huge blocks of data, say of the order of millions of points, can be stored in blocks of size (n+1)(n+1) where n is the usually small (typically less than 100) dimensional input space in which the data resides. To demonstrate the effectiveness of the algorithm we classify a dataset of 1 billion points in 10-dimensional input space into two classes in less than 2.5 hours on a 400 MHz Pentium II processor.

In (6(xii)), prior knowledge in the form of multiple polyhedral sets, each belonging to one of two categories, is introduced into a reformulation of a linear support vector machine classifier. The resulting formulation leads to a linear program that can be solved efficiently. Real world examples, from DNA sequencing and breast cancer prognosis, demonstrate the effectiveness of the proposed method. Numerical results show improvement in test set accuracy after the incorporation of prior knowledge into ordinary data-based linear support vector machine classifiers. One experiment also shows that a linear classifier, based solely on prior knowledge, far outperforms the direct application of the prior knowledge rules to classify new examples.

In (6(xiii)), Characterization of the containment of a polyhedral set in a closed halfspace, a key factor in generating knowledge-based support vector machine classifiers, is extended to the following:
(a) Containment of one polyhedral set in another.
(b) Containment of a polyhedral set in a reverse-convex set defined by convex quadratic constraints.
(c) Containment of a general closed convex set, defined by convex constraints, in a reverse-convex set defined by convex nonlinear constraints.

The first two characterizations can be determined in polynomial time by solving m linear programs for (a) and m convex quadratic programs for (b), where m is the number of constraints defining the containing set. In (c), m convex programs need to be solved in order to verify the characterization, where again m is the number of constraints defining the containing set. All polyhedral sets, like the knowledge sets of support vector machine classifiers, are characterized by the intersection of a finite number of closed halfspaces.

In (6(xiv)), a fundamental classification problem of data mining and machine learning is that of minimizing a strongly convex, piecewise quadratic function on the n-dimensional real space. We show finite termination of a Newton method to the unique global solution starting from any point in the space. If the function is well conditioned, then no stepsize is required from the start, and if not, an Armijo stepsize is used. In either case finite termination is guaranteed to the unique global minimum solution.

In (6(xv)), an implicit Lagrangian formulation of a support vector machine classifier, proposed earlier by one of the authors and which led to a highly effective iterative scheme, is solved here by a finite Newton method. The proposed method, which is extremely fast and terminates in 6 or 7 iterations, can handle classification problems in very high dimensional spaces, e.g. over 28,000, in a few seconds on a 400 MHz Pentium II machine. The method can also handle problems with large datasets and requires no specialized software other than a commonly available solver for a system of linear equations. Finite termination of the proposed method is established in this work.

In (6(xvi)), a fast Newton method is proposed for solving linear programs with a very large (approx 1 million) number of constraints and a moderate (approx 100) number of variables. Such linear programs occur in data mining and machine learning. The proposed method is based on the apparently overlooked fact that the dual of an asymptotic exterior penalty formulation of a linear program provides an EXACT least 2-norm solution to the dual of the linear program for FINITE values of the penalty parameter but NOT for the primal linear program. Solving the dual for a finite value of the penalty parameter yields an exact least 2-norm solution to the dual, but NOT a primal solution unless the parameter approaches zero. However, the exact least 2-norm solution to dual problem can be used to generate a highly accurate primal solution if m is much bigger than n and the primal solution is unique. Utilizing these facts, a fast globally convergent finite Newton method is proposed. A simple prototype of the method is given in eleven lines of MATLAB code. This code is capable of solving a primal linear program with two million constraints and a hundred variables in 17.4 minutes to an extremely high accuracy of on a 400Mhz Pentium II with 2 gigabytes of memory. CPLEX 6.5 ran out of memory in attempting to solve the same problem on the same machine.

In (6(xvii)), a fast Newton method, that suppresses input space features, is proposed for a linear programming formulation of support vector machine classifiers. The proposed stand-alone method can handle classification problems in very high dimensional spaces, such as 28,032 dimensions, and generates a classifier that depends on very few input features, such as 7 out of the original 28,032. The method can also handle problems with a large number of data points and requires no specialized linear programming packages but merely a linear equation solver. For nonlinear kernel classifiers, the method utilizes a minimal number of kernel functions in the classifier that it generates.

5. Personnel Supported

Yuh-Jye Lee, Ph.D. Candidate. Degree granted August 2001.
Glenn Fung, Ph.D. Candidate. Degree expected August 2003.
Michael Thompson, Ph.D. Candidate. Degree expected August 2005

6. Publications

(i) Y.-J. Lee, O. L. Mangasarian & W. H. Wolberg: ``Breast Cancer Survival and Chemotherapy: A Support Vector Machine Analysis", Data Mining Institute Technical Report 99-10, December 1999. DIMACS Series in Discrete Mathematics and Computer Science, Volume 5, American Mathematical Society 2000, 1-10.
ftp://ftp.cs.wisc.edu/pub/dmi/tech-reports/99-10.ps

(ii) G. Fung & O. L. Mangasarian: ``Data Selection for Support Vector Machine Classifiers", Data Mining Institute Technical Report 00-02, February 2000. "KDD-2000", Proceedings of the Sixth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, August 20-23, 2000, Boston, MA, R. Ramakrishnan & S. Stolfo, editors, ACM, NY 2000, 64-70.
ftp://ftp.cs.wisc.edu/pub/dmi/tech-reports/00-02.ps

(iii) O. L. Mangasarian & D. R. Musicant: ``Active Support Vector Machine Classification", Data Mining Institute Technical Report 00-04, April 2000. Neural Information Processing Systems 2000 (NIPS 2000), Todd K. Lee, Thomas G. Dietterich and Volker Tresp, editors, MIT Press 2001, 577-583.
ftp://ftp.cs.wisc.edu/pub/dmi/tech-reports/00-04.ps

(iv) O. L. Mangasarian & D. R. Musicant: ``Lagrangian Support Vector Machines", Data Mining Institute Technical Report 00-06, June 2000. Journal of Journal of Machine Learning Research 1, March 2001, 161-177.
ftp://ftp.cs.wisc.edu/pub/dmi/tech-reports/00-06.ps

(v)Y.-J. Lee, O. L. Mangasarian: "RSVM: Reduced Support Vector Machines", Data Mining Institute Technical Report 00-07, August 2000. CD Proceedings of the First SIAM International Conference on Data Mining, Chicago, April 5-7, 2001, SIAM, Philadelphia, ISBN 0-89871-495-8.
ftp://ftp.cs.wisc.edu/pub/dmi/tech-reports/00-07.ps

(vi) G. Fung, O. L. Mangasarian and A. J. Smola: ``Minimal Kernel Classifiers", Data Mining Institute Technical Report 00-08, November 2000.

Journal of Machine Learning Research 3, 2002, 303-321.
ftp://ftp.cs.wisc.edu/pub/dmi/tech-reports/00-08.ps

(vii) G. Fung and O. L. Mangasarian: ''Proximal Support Vector Machine Classifiers", Data Mining Institute Technical Report 01-02, February 2001. Proceedings KDD-2001: Seventh International Conference on Knowledge Discovery and Data Mining, San Francisco August 26-29, 2001. Association for Computing Machinery, New York, 2001, 77-86.
ftp://ftp.cs.wisc.edu/pub/dmi/tech-reports/01-02.ps

(viii) Y.-J. Lee, O. L. Mangasarian and W. H. Wolberg):''Survival-Time Classification of Breast Cancer Patients", Data Mining Institute Technical Report 01-03, March 2001. Data Mining Institute Technical Report 01-03, March 2001. Computational Optimization and Applications, to appear.
ftp://ftp.cs.wisc.edu/pub/dmi/tech-reports/01-03.ps

(ix) O. L. Mangasarian: ''Data Mining via Support Vector Machines", Data Mining Institute Technical Report 01-05, May 2001. IFIP Conference on System Modelling and Optimization Proceedings, Trier, Germany, July 23-27, 2001, to appear.
ftp://ftp.cs.wisc.edu/pub/dmi/tech-reports/01-05.ps

(x) G. Fung and O. L. Mangasarian: ''Multicategory Support Vector Machine Classifiers", Data Mining Institute Technical Report 01-06, July 2001. Machine Learning, submitted.
ftp://ftp.cs.wisc.edu/pub/dmi/tech-reports/01-06.ps

(xi) G. Fung and O. L. Mangasarian: "Incremental Support Vector Machine Classification", Data Mining Institute Technical Report 01-08, September 2001. Proceedings of the Second SIAM International Conference on Data Mining, Arlington, Virginia, April 11-13, 2002,R. Grossman, H. Mannila and R. Motwani (editors), SIAM, Philadelphia 2002, 247-260.
ftp://ftp.cs.wisc.edu/pub/dmi/tech-reports/01-08.ps

(xii) Glenn Fung, O. L. Mangasarian and Jude Shavlik: " Knowledge-Based Support Vector Machine Classifiers", Data Mining Institute Technical Report 01-09, November 2001. Neural Information Processing Systems 2002 (NIPS 2002), Vancouver, BC, December 10-12, 2002, to appear.
ftp://ftp.cs.wisc.edu/pub/dmi/tech-reports/01-09.ps

(xiii) O. L. Mangasarian: "Set Containment Characterization", Data Mining Institute Technical Report 01-10, November 2001. Journal of Journal of Global Optimization 24(4) December 2002, 473-480.
ftp://ftp.cs.wisc.edu/pub/dmi/tech-reports/01-10.ps

(xiv) O. L. Mangasarian: " A Finite Newton Method for Classification Problems", Data Mining Institute Technical Report 01-11, December 2001. Optimization Methods and Software 17, 2002, 913-929.
ftp://ftp.cs.wisc.edu/pub/dmi/tech-reports/01-11.ps

(xv) G. Fung and O. L. Mangasarian: "Finite Newton Method for Lagrangian Support Vector Machine Classification", Data Mining Institute Technical Report 02-01, February 2002. Nuero-Computing, submitted.
ftp://ftp.cs.wisc.edu/pub/dmi/tech-reports/01-11.ps

(xvi) O. L. Mangasarian: "A Newton Method for Linear Programming", Data Mining Institute Technical Report 02-02, March 2002. Mathematical Programming, submitted.
ftp://ftp.cs.wisc.edu/pub/dmi/tech-reports/02-02.ps

(xvii) G. Fung and O. L. Mangasarian: ``A Feature Selection Newton Method for Support Vector Machine Classification", Data Mining Institute Technical Report 02-03, September 2002. Computational Optimization and Applications, submitted.
ftp://ftp.cs.wisc.edu/pub/dmi/tech-reports/02-03.ps

7. Interactions/Transitions

a. Meetings, Conferences & Seminars

(I) Seminar, University of California at San Diego, January 11, 2000
Talk: "SSVM: A Smooth Support Vector Machine for Classification"

(II) INFORMS National Meeting, May 7-10, 2000, Salt Lake City
Talk: "Support Vector Regression"

(III) International Symposium on Mathematical Programming, August 7-11, Atlanta
Talk: "Support Vector Machines for Data Classification and Mining"
Talk: "Support Vector Machine Regression"

(IV) KDD00: International Conference on Knowledge Discovery & Data Mining, August 20-23, 2000, Boston
Talk: "Data Selection for Support Vector Machine Classification"

(V) MIT Distinguished Speaker Series, Cambridge, MA October 4, 2001
Talk:"Mathematical Programming in Support Vector Machines"
(Video CD of talk available from PI on request)

(VI) INFORMS Annual Meeting, San Antonio, TX, November 5-8, 2000
Talk: "Unlabeled Data Classification by Support Vector Machines"

(VII) Neural Information Processing Systems (NIPS-2000), Denver, CO, November 28-30, 2000
Talk: ``Active Support Vector Machine Classification"

(VIII) Neural Information Processing Systems (NIPS-2000), Workshop on New Perspectives in Kernel-Based Learning, Breckenridge, CO, December 1-2, 2000
Talk: "LSVM: Lagrangian Support Vector Machines"

(IX) Seminar, University of California at San Diego, January 9, 2001
Talk: "Support Vector Machines via Mathematical Programming"

(X) SIAM Data Mining Conference, Chicago, April 5-7, 2001
Talk: "RSVM: Reduced Support Vector Machines"

(XI) INRIA Rocquencourt, France, July 17, 2001

Talk: "Mathematical Programming for Support Vector Machine Classifiers"

(XII) IFIP 2001, Trier, Germany, July 23-27, 2001
Talk: "Data Mining via Support Vector Machines" (Opening Plenary)

(XIII) KDD2001 San Francisco, CA, August 26-29, 2001
Talk: "Proximal Support Vector Machine Classifiers"

(XIV) UW-Madison, December 2001
G. Fung, O. L. Mangasarian & J. Shavlik:
Talk: "Knowledge-Based Support Vector Machine Classifiers"

(XV) Second SIAM International Conference on Data Mining
SDM 2002 Arlington, Virginia, April 11-13, 2002
G. Fung & O. L. Mangasarian
Talk: "Incremental Support Vector Machine Classification"

(XVI) CSNA 2002: Classification
Society of North America Annual Meeting, Madison, Wisconsin, June 13-16, 2002
G. Fung, O. L. Mangasarian:
Talk: "The Disputed Federalist Papers: SVM Feature Selection via Concave Minimization"

(XVII) Mathematics Department, University of California, San Diego, July 26, 2002
Talk: "A Newton Method for Linear Programming"

8. New discoveries, inventions, or patent disclosures

Patent Application Number US 10/114,419 was filed on April 1, 2002, for the LSVM (Lagrangian Support Vector Machine) classifier, described in AFOSR-supported report: ftp://ftp.cs.wisc.edu/pub/dmi/tech-reports/00-06.ps
SAS, a major software company has signed up for the exclusive rights for this pending patent.

9. Honors/Awards

INFORMS Lanchester Prize "awarded annually for the best contribution, published in English, to the field of operations research and management sciences", was awarded to Olvi Mangasarian at the INFORMS annual meeting, San Antonio, TX, November 5-8, 2000. See:
http://www.informs.org/Prizes/LanchesterDetails.html
for more details.